

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-306644

(43)公開日 平成11年(1999)11月5日

(51)Int.Cl. ⁸	識別記号	F I
G 1 1 B 19/02	5 0 1	G 1 1 B 19/02 5 0 1 S
G 0 6 F 3/06	3 0 6	G 0 6 F 3/06 3 0 6 B
	5 4 0	5 4 0
G 1 1 B 19/04	5 0 1	G 1 1 B 19/04 5 0 1 D
20/10		20/10 D

審査請求 未請求 請求項の数3 OL (全 6 頁)

(21)出願番号 特願平10-107461

(22)出願日 平成10年(1998)4月17日

(71)出願人 000232047

日本電気エンジニアリング株式会社
東京都港区芝浦三丁目18番21号

(72)発明者 尾原 寛子

東京都港区芝浦三丁目18番21号 日本電気
エンジニアリング株式会社内

(72)発明者 市川 文男

東京都港区芝浦三丁目18番21号 日本電気
エンジニアリング株式会社内

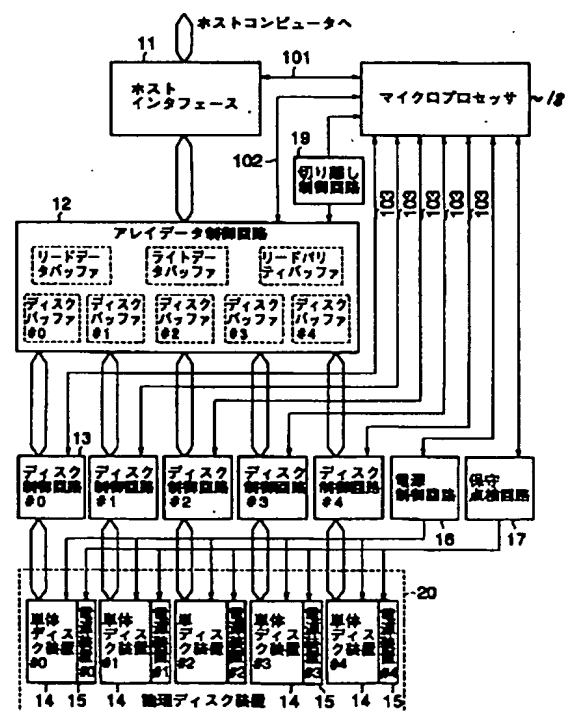
(74)代理人 弁理士 京本 直樹 (外2名)

(54)【発明の名称】 ディスクアレイ装置

(57)【要約】

【課題】 故障が発生したディスク装置に対して、故障診断を行うことができるディスクアレイ装置を提供する。

【解決手段】 複数の単体ディスク装置14と、これら単体ディスク装置とホストコンピュータとの間に接続され、複数の単体ディスク装置が単一の論理ディスク装置として動作するように制御するマイクロプロセッサ18を有するディスクアレイ装置において、マイクロプロセッサが複数の単体ディスク装置を監視して故障した単体ディスク装置を検出したとき、故障した単体ディスク装置をリード動作及びライト動作に関して論理ディスク装置から切り離す切り離し制御回路19と、マイクロプロセッサからの指示に従って、故障した単体ディスク装置の故障診断を行う点検保守回路とを設けた。



【特許請求の範囲】

【請求項1】 複数の単体ディスク装置と、該複数の単体ディスク装置とホストコンピュータとの間に接続され、前記複数の単体ディスク装置が単一の論理ディスク装置として動作するようにこれら複数の単体ディスク装置を制御する制御手段とを有するディスクアレイ装置において、前記制御手段が前記複数の単体ディスク装置を監視して故障した単体ディスク装置を検出したとき、該故障した単体ディスク装置をリード動作及びライト動作に関して前記論理ディスク装置から切り離す切り離し手段と、前記制御手段と前記複数の単体ディスク装置とに接続され、前記制御手段からの指示に従って、前記故障した単体ディスク装置の故障診断を行う故障診断手段とを備えたことを特徴とするディスクアレイ装置。

【請求項2】 前記制御手段と前記複数の単体ディスク装置とに接続され、前記複数の単体ディスク装置の電源のオン/オフを制御する電源制御手段を有することを特徴とする請求項1のディスクアレイ装置。

【請求項3】 複数のディスク装置と、該複数のディスク装置にそれぞれ接続され各ディスク装置に対するデータの書き込み及び読出しを行う複数のディスク制御装置と、該複数のディスク装置に接続され論理アドレスと物理アドレスとの変換を行うとともに書き込みデータの分解と読出しデータの統合とを行なう前記複数のディスク装置を1台の論理ディスク装置と見なすアレイデータ制御回路と、該アレイデータ制御回路をホストコンピュータに接続するためのインタフェース回路と、該インタフェース回路、前記アレイデータ制御回路、及び前記ディスク制御回路に接続され、これらを制御するとともに、前記ディスク制御回路を介して前記複数のディスク装置を監視し、前記データの書き込み及び読出しを実行させるマイクロプロセッサとを有するディスクアレイ装置において、前記マイクロプロセッサが前記複数のディスク装置に故障を検出したとき出力する、故障したディスク装置を切り離すための切り離し情報を保持し、前記アレイデータ制御回路へ供給する切り離し制御回路と、前記マイクロプロセッサからの要求に従い、特定のディスク装置の診断を行う保守回路とを設け、前記アレイデータ制御回路は前記切り離し制御回路からの前記切り離し情報に基づいてリード動作及びライト動作に関して故障したディスク装置を前記論理ディスク装置から切り離すとともに、切り離されたディスク装置の診断を行うようにしたことを特徴とするディスクアレイ装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、ディスクアレイ装置に関する。

【0002】

【従来の技術】 ディスクアレイ装置は、小型で安価なディスク装置を複数台並列接続して、1台の論理ディスク

装置として利用できる様にした装置である。このようなディスクアレイ装置の制御方式として、“A Case for Redundant Array of Inexpensive Disks (=RAID)” (Technical Report UCB/CSD 87/391, December 1987) において、RAID-1~5が提案されている。

【0003】 例えば、RAID-3は、複数のディスク全てに対して同時にアクセスすることにより、高い転送レートを実現する方式で、スーパーコンピュータ等、一度に大量のデータを処理する装置に適している。ところが、RAID-3は、小容量のデータを取り扱う際にも、全てのディスクにアクセスするため、効率が悪い。そこで、RAID-5では、パリティデータを各ディスクに分散させて、リード/ライト処理の多重化を目指している。最近、トランザクションアプリケーションをターゲットとしたRAID-5ディスクアレイ装置が、各社から発表されている。

【0004】 これらのRAID-1~5のディスクアレイ装置では、複数のディスク装置のいずれかに故障が発生しても、残りのディスク装置を用いてデータの復元ができるように構成されているので、故障したディスク装置をできるだけ早く切り離し、その影響を取り除くように制御される。

【0005】

【発明が解決しようとする課題】 従来のディスクアレイ装置は、複数のディスク装置のいずれかに故障が発生すると、故障が発生したディスク装置を直ちに切り離すよう制御される。このため、故障の原因を検出することができないという問題点がある。また、故障したディスク装置に代えて新たに接続されたディスク装置が、正常に動作するかどうかは、実際に、交換されたディスク装置にアクセスしなければ、分からないという問題点がある。

【0006】 本発明は、故障が発生したディスク装置に対して、通常のリード/ライト動作を停止するとともに、その故障診断を行うことができるディスクアレイ装置を提供することを目的とする。また、故障したディスク装置に代えて接続されたディスク装置に対しても診断を行うことができるディスクアレイ装置を提供することを目的とする。

【0007】

【課題を解決するための手段】 本発明によれば、複数の単体ディスク装置と、該複数の単体ディスク装置とホストコンピュータとの間に接続され、前記複数の単体ディスク装置が単一の論理ディスク装置として動作するようにこれら複数の単体ディスク装置を制御する制御手段とを有するディスクアレイ装置において、前記制御手段が前記複数の単体ディスク装置を監視して故障した単体ディスク装置を検出したとき、該故障した単体ディスク装置をリード動作及びライト動作に関して前記論理ディスク装置から切り離す切り離し手段と、前記制御手段と前

記複数の単体ディスク装置とに接続され、前記制御手段からの指示に従って、前記故障した単体ディスク装置の故障診断を行う故障診断手段とを備えたことを特徴とするディスクアレイ装置が得られる。

【0008】また本発明によれば、複数のディスク装置と、該複数のディスク装置にそれぞれ接続され各ディスク装置に対するデータの書き込み及び読出しを行う複数のディスク制御装置と、該複数のディスク装置に接続され論理アドレスと物理アドレスとの変換を行うとともに書き込みデータの分解と読出しデータの統合を行なう前記複数のディスク装置を1台の論理ディスク装置と見なすアレイデータ制御回路と、該アレイデータ制御回路をホストコンピュータに接続するためのインタフェース回路と、該インタフェース回路、前記アレイデータ制御回路、及び前記ディスク制御回路に接続され、これらを制御するとともに、前記ディスク制御回路を介して前記複数のディスク装置を監視し、前記データの書き込み及び読出しを実行させるマイクロプロセッサとを有するディスクアレイ装置において、前記マイクロプロセッサが前記複数のディスク装置に故障を検出したとき出力する、故障したディスク装置を切り離すための切り離し情報を保持し、前記アレイデータ制御回路へ供給する切り離し制御回路と、前記マイクロプロセッサからの要求に従い、特定のディスク装置の診断を行う保守回路とを設け、前記アレイデータ制御回路は前記切り離し制御回路からの前記切り離し情報に基づいてリード動作及びライト動作に関して故障したディスク装置を前記論理ディスク装置から切り離すとともに、切り離されたディスク装置の診断を行うようにしたことを特徴とするディスクアレイ装置が得られる。

【0009】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態について説明する。

【0010】図1に本発明の一実施の形態を示す。本実施の形態によるディスクアレイ装置は、ホストコンピュータ（図示せず）に接続されるホストインタフェース11と、ホストインタフェース11に接続されるアレイデータ制御回路12と、アレイデータ制御回路12に接続される複数のディスク制御回路13と、ディスク制御回路13にそれぞれ接続される単体ディスク装置14と、各単体ディスク装置14に接続される電源装置15と、電源装置15を制御する電源制御回路16と、単体ディスク装置14の保守点検を行う保守点検回路17と、ホストインタフェース11、アレイデータ制御回路12、ディスク制御回路13、電源制御回路16、及び保守点検回路17を制御するマイクロプロセッサ18と、マイクロプロセッサ18とアレイデータ制御回路12との間に接続された切り離し制御回路19とを有している。なお、本実施の形態では、5台の単体ディスク装置14（4台のデータディスク装置とパリティディスク装置）

で1台の論理ディスク装置20を構成するRAID-5ディスクアレイ装置を想定している。

【0011】次に、このディスクアレイ装置の動作について説明する。まず、リード動作について説明する。ホストコンピュータから出力されたデータリード命令は、ホストインタフェース11に入力される。マイクロプロセッサ18は、ホストインタフェース制御線101を介してホストインタフェース11を監視しており、ホストインタフェース11にデータリード命令が入力されたことを検出する。そして、マイクロプロセッサ18は、ホストインタフェース11から、読出スタート論理アドレス及び読出データセクタ数を受け取る。

【0012】マイクロプロセッサ18は、読出スタート論理アドレス及び読出データセクタ数を受け取ると、これらに基づいて、読み出そうとするデータを記憶している単体ディスク装置14の装置番号と、読出スタート物理アドレスとを求める。そして、求めた装置番号の単体ディスク装置14とアレイデータ制御回路12との間にリードデータパスを確立するよう、アレイデータ制御回路制御線102を介してアレイデータ制御回路12に指示を出す。

【0013】次に、マイクロプロセッサ18は、先に求めた装置番号を有する単体ディスク装置14へのアクセスが可能かどうか、ディスクインタフェース制御線103を介してディスク制御回路13から送られてくるステータス情報に基づいて判断する。アクセスが可能であれば、マイクロプロセッサ18は、ディスクインタフェース制御線103を介してそのディスク制御回路13へ、読出スタート物理アドレス及び読出データセクタ数を与える。

【0014】ディスク制御回路13は、マイクロプロセッサ18から読出スタート物理アドレス及び読出データセクタ数が与えられると、マイクロプロセッサ18に対してビジーを示すステータス情報を出力する。このステータス情報の出力は、リード動作が終了するまで続く。また、ディスク制御回路13は、マイクロプロセッサ18から与えられた読出スタート物理アドレス及び読出データセクタ数に従って、単体ディスク装置14からデータを読出し、アレイデータ制御回路12へ出力する。アレイデータ制御回路12は、読み出されたデータを一旦ディスクバッファへ格納し、さらにリードデータバッファに格納しなおして、ホストインタフェース11を介してホストコンピュータへ読出したデータを出力する。ディスク制御回路13は、ホストコンピュータへのデータ送出が終了すると、ステータス情報をアクセス可能に変更してリード動作を終了する。なお、データの読出しが複数の単体ディスク装置14に対して行われる場合には、各ディスクバッファに格納された読出データをリードデータバッファに格納し直す際に、一つのデータとして組み立てが行われる。こうして、ホストコンピュータ

からのデータリード命令に回答して、所望のデータが読み出される。

【0015】次に、このディスクアレイ装置のライト動作について説明する。このディスクアレイ装置では、4台の単体ディスク装置14（データディスク装置）に記憶されているデータからパリティデータを生成して、もう1台の単体ディスク装置14（パリティディスク装置）に記憶させ、いずれかの単体ディスク装置14に障害が発生した場合に、その単体ディスク装置14に記憶されているデータを他の単体ディスク装置14に記憶されているデータから復元できるようにしている。このため、いずれかのデータディスク装置に対してライト動作を行う場合には、パリティデータの変更も必要となる。

【0016】パリティデータを変更する方法としては、書き込もうとするデータと、ライト動作の対象となるデータディスク装置以外のデータディスク装置から読出したデータをから新たなパリティデータを生成する方法や、データを書き込もうとするアドレスに存在するデータ（旧データ）と、書き込もうとするデータ（新データ）と、旧データに対応するパリティデータ（旧パリティデータ）とから、新たなパリティデータ（新パリティデータ）を生成する方法などがある。いずれにしても、ディスクアレイ装置へのライト動作を行う場合には、単体ディスク装置14からのリード動作を伴うことになる。以下では、旧データと新データと旧パリティデータとから新パリティデータを生成する方法を採用したライト動作について説明する。

【0017】まず、ホストコンピュータからのデータライト命令は、ホストインタフェース11に入力される。マイクロプロセッサ18は、ホストインタフェース11にデータライト命令が入力されたことを検出して、書込スタート論理アドレス及び書込データセクタ数を受け取る。そして、マイクロプロセッサ18は、書込スタート論理アドレス及び書込データセクタ数に基づいて、データを書き込むべき単体ディスク装置14の装置番号及び書込スタート物理アドレスを求める。ここで求められた物理アドレスは、旧データの読出スタート物理アドレスとして使用されるとともに、旧パリティデータの読出スタート物理アドレスとしても使用される。（パリティデータを複数の単体ディスク装置に分散させたディスクアレイ装置の場合には、データを書き込む領域に対応するパリティデータが記憶されている単体ディスク装置14の装置番号及びその領域の物理アドレスを同時に求めなくてはならない。）この後、マイクロプロセッサ18は、旧データと旧パリティデータとを読み出すために、求めた装置番号の単体ディスク装置14とアレイデータ制御回路12との間、及びパリティデータを記憶する単体ディスク装置14とアレイデータ制御回路12との間にリードデータバスを確立するようアレイデータ制御回路12に指示を出す。

【0018】次に、マイクロプロセッサ18は、求めた装置番号の単体ディスク装置14に接続されたディスク制御回路13からのステータス情報に基づいて、求めた装置番号を有する単体ディスク装置14へのアクセスが可能かどうか判断する。また、パリティデータを記憶する単体ディスク装置14に対しても同様にしてアクセスが可能かどうか判断する。そして、これらの単体ディスク装置14がアクセス可能であれば、マイクロプロセッサ18は、書込スタート物理アドレス（読出スタート物理アドレス）とデータセクタ数とをこれら単体ディスク装置14に接続されたディスクインタフェース13に供給する。

【0019】書込スタート物理アドレスとデータセクタ数とが供給されたディスク制御回路13は、マイクロプロセッサ18へビジーを示すステータス情報を出力し、単体ディスク装置14からそれぞれ旧データと旧パリティデータを読み出す。読み出された旧データは、アレイデータ制御回路12のディスクバッファを介してリードデータバッファに格納される。また、読み出された旧パリティデータは、ディスクバッファを介してリードパリティバッファに格納される。

【0020】旧データ及び旧パリティデータのリード動作が終了すると、ディスク制御回路13からマイクロプロセッサ18へ出力されるステータス情報がアクセス可能に変化する。このステータス情報の変化を受けてマイクロプロセッサ18は、アレイデータ制御回路12とホストインタフェース11との間にライトデータバスの確立を指示する。そして、マイクロプロセッサ18は、ホストインタフェース11に対してホストコンピュータからのライトデータ（新データ）の受け取りを指示する。ホストインタフェース11は、受け取った新データがアレイデータ制御回路12のライトデータバッファに転送されたことを確認して、転送終了をマイクロプロセッサ18に通知する。

【0021】上記の動作により、アレイデータ制御回路12のリードデータバッファ、ライトデータバッファ、及びリードパリティバッファには、それぞれ旧データ、新データ、及び旧パリティデータが格納されたことになる。アレイデータ制御回路12は、これら旧データ、新データ、及び旧パリティデータから新パリティデータを生成する。具体的には、これら旧データ、新データ、及び旧パリティデータの排他的論理和を求めれば、その和が新パリティデータとなる。詳述すると、新データと旧データとの排他的論理和は、新データと旧データとが異なる場合に“1”となり、新データと旧データと同じ場合に“0”となる。

【0022】次に、新パリティデータを単体ディスク装置に書き込む動作について説明する。

【0023】まず、マイクロプロセッサ18は、旧データの読み取りにおいて使用した書き込みスタートアドレ

10

20

30

40

50

ス及び書き込みデータセクタ数から、新データを書き込む単体ディスク装置番号と単体ディスク装置アドレスを求める。そして、マイクロプロセッサ18は、求めた単体ディスク装置番号に基づいてディスク制御回路13から制御信号線103を介して送られてくるステータス情報により、その単体ディスク装置14がアクセス可能か判断する。

【0024】アクセス可能であると判断した場合、マイクロプロセッサ18は、制御線103を介してディスク制御回路13に書き込み単体ディスク装置アドレスを転送する。ディスク制御回路13は、送られてきた書き込み単体ディスク装置アドレスに基づいて、単体ディスク装置14への書き込み動作を行う。この動作中に、ディスク制御回路13に対してマイクロプロセッサ18から命令が送られてきた場合には、その命令を受け付けられないという意味のステータス情報が、制御線103を介してマイクロプロセッサに送られる。

【0025】単体ディスク装置14が、書き込み動作を終了すると、ディスク制御回路13は、制御信号線103を介し、マイクロプロセッサ18へアクセス可能という意味の制御情報を送り、新パリティデータを単体ディスク装置に書き込む動作を終了する。

【0026】以上のようにして、単体ディスク装置への書き込み動作が行われる。

【0027】次に、本発明の特徴である、5台の単体ディスク装置14のうちの1台が故障した場合の単体ディスク装置の切り離し、及び診断方式について説明する。

【0028】本発明のディスクアレイ装置は、5台の単体ディスク装置14のうちの任意の1台を他の4台の単体ディスク装置14から切り離して制御する場合の切り離し情報を貯蔵すると同時に、アレイデータ制御回路12にその切り離し情報を提供する切り離し制御回路19と、N台の各単体ディスク装置14に電源を個別に供給する5台の電源装置15と、マイクロプロセッサ18の指示に基づいて電源の投入切断を実施する電源制御回路16と、マイクロプロセッサ18の指示に基づいて単体ディスク装置の診断を実施する保守点検回路17とを備えており、マイクロプロセッサ18は、単体ディスク装置14の制御中に異常を検出し、異常を検出した1台の単体ディスク装置14を切り離すべきだと判断した場合には、現在実行中の処理が終了した時点で、切り離し制

御回路19に対して、切り離すべき単体ディスク装置番号をセットする。この切り離すべき単体ディスク装置番号がセットされた後のアレイデータ制御回路12における読み出し及び書き込み命令処理は、切り離し単体ディスク装置14からのデータを無視し、残りの4台のディスクバッファのデータから切り離されたデータを復元しながら実行される。このとき同時に、切り離された単体ディスク装置に対して、マイクロプロセッサ18から保守点検回路17及び電源制御回路16を通じて、その単体ディスク装置14の故障原因究明のために、電源投入切断を含む故障診断を実行する。また、故障した単体ディスク装置14の交換後に新規単体ディスク装置14に対しても、同様の診断を実施するようにしてもよい。さらに、予防保守の一環として1台の単体ディスク装置14を他の4台のディスク装置から切り離して診断することも可能である。

【0029】

【発明の効果】以上説明したように、本発明によれば、1台の単体ディスク装置が故障した場合に、根故障した1台の単体ディスク装置を他の単体ディスク装置から切り離して制御すると同時に、きり話し中の故障した単体ディスク装置に対して、保守点検回路及び電源制御回路を用いた診断を実施し、単体ディスク装置の異常箇所を早期に発見するとともに、交換後の単体ディスク装置の診断や、予防保守として1台の単体ディスク装置を切り離しての診断を実施できる信頼性の高いディスクアレイ装置を提供することができる。

【図面の簡単な説明】

【図1】本発明の一実施の形態を示すブロック図である。

【符号の説明】

11	ホストインタフェース
12	アレイデータ制御回路
13	ディスク制御回路
14	単体ディスク装置
15	電源装置
16	電源制御回路
17	保守点検回路
18	マイクロプロセッサ
19	切り離し制御回路

【図1】

